DOI: https://doi.org/10.61762/pjstvol8iss2art004

PLATFORM - A Journal of Science & Technology

Received: 22 October 2025, Revised: 17 November 2025, Accepted: 23 November 2025, Published: 31 December 2025, Publisher: UTP Press, Creative Commons: CC BY 4.0

A ROBUST DEEP LEARNING MODEL FOR MONITORING SUBSURFACE USING DAS DATA FOCUSING ON NOISE, INTERPRETABILITY AND UNCERTAINTY

Nimra Iqbal[1]*, Izzatdin Abdul Aziz[1], Halimaton Saadiah Hakimi[2]

[1]Department of Computer and Information Sciences, Universiti Teknologi PETRONAS, Malaysia
[2]Department of Computing, Universiti Teknologi PETRONAS, Malaysia

*Corresponding author: nimra_24005120@utp.edu.my

ABSTRACT

Distributed acoustic sensing (DAS) is characterised by high-dimensional, temporal characteristics and high-frequency, continuous data, achieved by converting a fibre-optic cable into a nodal array of sensors. However, DAS data are often affected by noise issues and missing values caused by sensor failure, hardware failure, or environmental disturbances. Many researchers have applied the conventional long short-term memory (LSTM) method due to its ability to process long sequences of data; however, it has struggled with complex, high-spatial data, limited interpretation, and the inability to quantify uncertainty in predictions. Additionally, LSTMs are vulnerable to noisy inputs, which compromises their robustness and trustworthiness in real-time monitoring applications, such as early warning detection systems. To address these limitations, this paper examines the application of bidirectional long short-term memory (Bi-LSTM) networks with an attention mechanism for seismic event detection using DAS data. Due to the high dimensionality and complexity of DAS signals, deep learning models are prone to overfitting, which compromises their performance on unseen data. This study proposes an advanced deep learning framework incorporating regularisation techniques, including dropout, L2 regularisation, and the early stopping method. Dropout mitigates over-reliance on specific features by randomly deactivating neurons during training, while L2 regularisation penalises large weights to promote simpler, more generalisable models, and early stopping prevents overfitting by monitoring validation performance. Furthermore, Bayesian inference is integrated with Monte Carlo dropout to enable uncertainty estimation, allowing the model to provide confidence intervals for detected seismic activity. By combining strong regularisation strategies with uncertainty quantification, the proposed Bi-LSTM framework achieves a balance between accuracy and generalisation, demonstrating its potential as a reliable tool for real-time seismic event detection using complex DAS data.

Keywords: *Deep learning for seismology, non-stationary seismic signal, time-series analysis, earthquake detection, supervised learning algorithm*

INTRODUCTION

Early warning systems play a vital role in detecting real-time seismic events and help to prevent the risk of disaster [1]. The use of traditional methods, including the empirical method, theoretical method, and manual interpretation, is limited to assessing the efficiency and robustness of the model [2]. When the model deals with large, high-dimensional, complex data generated by distributed acoustic sensing (DAS) [3]. DAS follows the fiber optics mechanism and Rayleigh scattering when light travel inside the glass tube of fiber optics, a small part of light are revert and scatter in all direction, the sensing mechanism of DAS is that external vibration such as seismic event or any machinery sound, traffic, vibrioses sound, and footstep are arise, they causes small changes in the properties of fiber cable such as displacement(movement of object near the fiber), refractive index, acoustic pressure (high intensity of sound waves), strain (vibration of stretching of cable along the fiber), temperature (DAS is responsive to change of temperature), distributed temperature

sensing (DTS) technology are embedded with DAS to monitor the temperature [4]. The small changes in strain and temperature properties of DAS can affect the phase shift due to vibration, change of phase, and intensity due to small temperature changes in the Rayleigh light backscattered phenomena. However, DAS data have some limitations, including high dimensionality, excessive noise, and redundant fields [5]-[6].

The long short-term memory (LSTM) approach, a type of recurrent neural network (RNN), is capable of capturing long-duration sequential data dependencies. Previous studies and researchers, such as [7] and [8], have made significant contributions to the development of LSTM and CNN for detecting seismic events. However, their models struggle with overfitting of data, sensitivity to noise, and poor generalisation of the model.

This paper introduces an advanced deep learning framework that combines bidirectional long short-term memory (Bi-LSTM) networks, attention mechanisms, and a denoising autoencoder.

LITERATURE REVIEW

This research presents an advanced machine learning framework integrating Bi-LSTM, attention mechanisms, denoising autoencoders, dropout, early stopping, and weight regularisation (L1/L2). This advanced model enhances the performance, interpretability, and robustness of the model, making it more efficient for detecting seismic events in DAS datasets [9]. The increasing accessibility of DAS data presents both opportunities and challenges for seismology researchers. Traditional seismic detection methods and analysis of seismic events struggle to cope with the large volume and high complexity of data. A deep learning approach proposed an advanced automated framework that introduces a feature extraction method and interpretable pattern recognition in raw waveform data [10].

RNNs of the type LSTM develop models to work with sequential data, allowing for long-term dependencies [11]. The model and idea proposed by [12] have a problem with RNNs, specifically that they forget the patterns of earlier sequences for long-term dependencies. The model struggles to learn long sequences and becomes increasingly smaller or vanishes for long sequential data. LSTMs address the issue of the vanishing gradient problem presented in existing RNNs, ensuring the ability to capture long sequential and temporal patterns in large, extended sequences of data [13].

DAS has revolutionised the monitoring of seismic activity by enabling researchers to collect detailed, continuous data over large areas. This provides very fine-grained information that is useful for detecting leaks of carbon dioxide and monitoring underground structures [14]. However, working with DAS data can be challenging because it is highly complex, has numerous dimensions, and contains significant noise. Traditional methods, such as unidirectional LSTM, work well for simple sequences but struggle to handle the complex time patterns and detailed space-time connections found in seismic DAS data [8],[15]. Additionally, standard LSTM models are not very transparent about their decision-making process, struggle to handle noise, and cannot indicate the certainty of their predictions, which is crucial for tasks such as monitoring carbon storage.

To address these issues, recent studies have shown that using Bi-LSTM is beneficial for modelling seismic signals, particularly when there are long and complex time patterns. Bi-LSTM examines data in both forward and backward directions, which enables the model to comprehend the full context and more accurately represent how seismic waves propagate through complex environments [16]. Adding attention mechanisms to Bi-LSTM also improves performance by allowing the model to focus on the most important parts of the data and time points, which helps it identify weak seismic signals in noisy data. Li et al. [17] found that incorporating attention into Bi-LSTM makes the model more comprehensible and improves accuracy in tasks such as classifying seismic signals. More recently, Zhu et al. [18] developed a multi-scale attention-Bi-LSTM framework that extracts different levels of temporal features from noisy data, resulting in improved performance across various space-time domains. These techniques are beneficial for DAS, where different frequencies and time ranges can have different geological meanings. Hybrid models that combine Bi-LSTM with attention mechanisms have shown great potential for detecting seismic events [19]. Bi-LSTM helps understand the time-based changes in signals. This mix is highly effective in handling noisy and complex data from the Earth, and it develops a

PLATFORM VOLUME 8 NUMBER 2 2025 e-ISSN: 26370530 42

hybrid model that utilises CNNs, Bi-LSTM, and attention for real-time seismic detection [20]. This model improved both the accuracy and the reliability of the detection. These types of models work well with the ConvNetQuake dataset, where CNNs can be combined with Bi-LSTM's ability to remember information in both directions for better learning over time [21].

One important aspect of utilising machine learning for monitoring CO_2 is understanding the uncertainty associated with the predictions, particularly when these predictions are used to inform decisions regarding safety or regulations. In this area, Bayesian deep learning approaches, such as MC dropout, offer a practical way to measure uncertainty in deep neural networks [13]. Nicolis et al. [22] introduced MC dropout as a method for estimating uncertainty without altering the model's training process. Lomax et al. [23] later demonstrated that this technique can distinguish between two types of uncertainty, those arising from the model's knowledge and those inherent to the data. Recent work, such as the study by Vaswani et al. [15], has highlighted the growing need for models that can understand and handle uncertainty when monitoring the Earth's subsurface using fibre-optic sensors. Effective pre-processing is also essential when dealing with raw DAS data, as it is often corrupted by environmental and system-induced noise [24]. Signal denoising techniques such as the Butterworth filter, wavelet transform, and standard scaling have been widely adopted to enhance signal quality before training ML models [25]. Jennifer [26] demonstrated the benefits of wavelet-based filtering in isolating meaningful seismic events from noisy DAS data, resulting in improved feature extraction and enhanced downstream model performance. A standard scaler, on the other hand, ensures uniform scaling of input features, helping recurrent models, such as Bi-LSTM, converge faster and avoid bias due to varying signal amplitudes [27]-[28].

In summary, this research proposes an architecture comprising a Bi-LSTM with attention and Bayesian inference through MC dropout, which addresses the core limitations of traditional LSTM models by offering bidirectional temporal modelling, adaptive attention, and predictive uncertainty estimation, all while being enhanced by robust pre-processing techniques. This integration of cutting-edge techniques aligns with the most recent advancements in seismic ML modelling. It contributes meaningfully to the ongoing effort of improving reliability, interpretability, and trustworthiness in DAS-based CO_2 monitoring systems.

Bidirectional LSTM Architecture

A bidirectional LSTM is an enhanced version of the LSTM that operates on long input sequences in both forward and backward directions, utilising two separate layers of LSTM. This allows the model to save the sequence from the past, which is the left side of the architecture and future, which is the right side of the architecture, concurrently, particularly useful for the task of long sequential labelling, learning of temporal patterns, speech recognition, early warning systems and natural language processing (NLP) where full cover of this network will improve accuracy [29]. The outcome is a merge of both forward and backward LSTM outputs at each time stamp, leading to high-resolution feature extraction [30].

Attention Mechanism

The attention mechanism is a method in deep learning networks to work on the important features of the input series when producing the output. Despite converting the entire input into a fixed-length format, similar to traditional LSTMs or RNNs, the attention mechanism allows the model to weigh separate input elements for every outcome step. It learns what to receive depending on the context, enhancing performance in tasks such as machine learning, text summarisation, and event detection for time-series data. This mechanism helps to capture long-term dependencies, particularly those involving long or complex sequences, and ensures the model is both efficient and accurate [31]-[32]. The model enhances reliability, and the learned attention weights reveal the impact of input elements on a specific prediction. When integrated with architectural elements, such as Bi-LSTM and attention mechanismsthe model's capability to understand and process sequential data was enhanced [33].

Bayesian Inference

The Bayesian approach is a statistical and probabilistic method for updating the likelihood of a hypothesis as more evidence or data become available. The Bayesian inference helps to ensure that predictions come out from the LSTM, such as the detection of seismic events with a measure of confidence scores that indicate how

sure the model is about their predictions, which is critical for making knowledgeable and safe decisions in crucial systems like early warning earthquake systems and CO_2 monitoring[34]-[35]—the idea of Bayes' Theorem. It combines prior belief (prior probability) before observing the data with new data (likelihood) to provide feedback on how likely the data is to be seen by the model. It modifies the belief after observing the data (posterior update) [36].

METHODS

Figure 1 illustrates a conceptual block that represents a significant component of the proposed system. It is applied to graphically present the methodology and emphasise the framework structure without presenting low-level details. The use of this kind of representation facilitates a clear flow of the system when making presentations and evaluations.

Dataset Preparation

The dataset related to code and problem will be compiled from publicly accessible websites, including GitHub repositories, the Incorporated Research Institutions for Seismology (IRIS) for seismic data, the United States Geological Survey (USGS) for earthquake events, Kaggle, and the Stanford Earthquake Dataset (STEAD). Data pre-processing will be performed to remove missing values and noisy data by applying filters such as a bandpass filter and an autoencoder, and by windowing the data

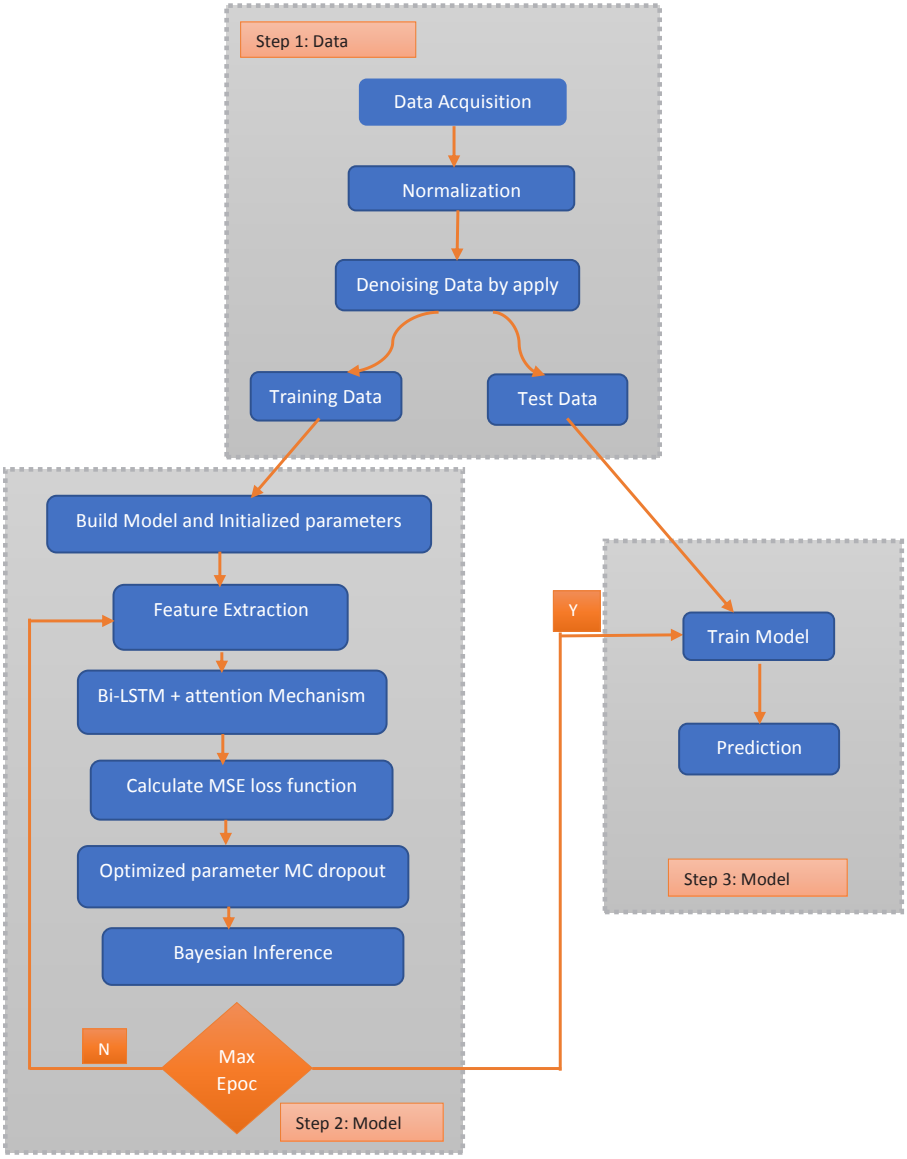

Figure 1 Research methodology of the proposed model

into small chunks for sequence learning. SMOTE is used to balance the data and avoid the problems of underfitting and convergence. Splitting the dataset into train and test sets for further inspection.

Model Selection

In this research, the proposed architectures integrate the Bi-LSTM for finding forward and backward dependencies in long sequential seismic data, along with an attention mechanism that focuses on the most important and relevant features in the output by assigning weights separately. The integration of Bayesian inference and Monte Carlo dropout (MC dropout) ensures uncertainty prediction in the output, reducing the overfitting issue commonly present in conventional LSTM models, and enables stochastic prediction under data uncertainty. This proposed architecture enables the model to develop more robust, trustworthy, and interpretable features.

The proposed architectural design is implemented using a library of Python, such as PyTorch and TensorFlow. MC dropout is imposed in both training and testing phases to produce a distribution of predictions. Bayesian inference is concatenated with multiple MC dropout to compute the mean (final output) and variance (confidence score).

Pseudocode (Short version)

```
Input: Raw DAS/Seismic signals I

Output: Event label (ŷ) and attention based explanation A'

I ← Filter → WaveletDenoise → Normalize

Z ← Windowize(I)

P ← BiLSTM → BiLSTM

A' ← Attention(P)

W ← AttentionWeightedSum(P, A')

if MC_dropout: W ← MC_Drop(W)

ŷ ← Softmax(W)

Train(ŷ, y)

Inference

(ŷ,A') ←model.predicted(x)

If MC_Dropout enabled then

Compute predictive mean (μ) and uncertainity (σ)

Return predictions + attention + uncertainty
```

Model Evaluation

The model will be evaluated using benchmark datasets such as STEAD. The efficiency of the proposed framework will be evaluated based on the following performance key metrics, including:

1. F1 score
2. Precision
3. Recall
4. Predictive uncertainty, such as the mean and covariance
5. Root Mean Square Error (RMSE)
6. Event Detection Latency
7. Area Under Curve (AUC)
8. Mean Absolute Error (MAE)
9. Probability Coverage

Compare with the Baseline Model

The performance of the Bi-LSTM will be compared with that of the baseline model's conventional LSTM, including unidirectional and GRU models. This comparison highlights differences in performance across the defined metrics. The final test is cross-validation against different time windows and seismic events. Analysing attention weight and interval of prediction to validate trustworthiness and interpretability of the model. This proposed methodology not only enhances the model's efficacy but also ensures reliability and transparency, meeting the requirements of a practical system in real-world seismic events.

RESULTS AND DISCUSSIONS

In Figure 2, the seismic data contains a 3-channel seismic waveform for noise and events. The dataset is uploaded in Python using the HDF5 format. There are 1,198 validated noise samples and 1,773 validated events, with each sample having a shape of 1,001 timesteps and three channels (Z, N, and E) (MN-AQU-BH). MN is for the Network code, AQU is for the station name, and Broadband high-gain channel. Channel 1 tells one of three components recorded, such as vertical, north-

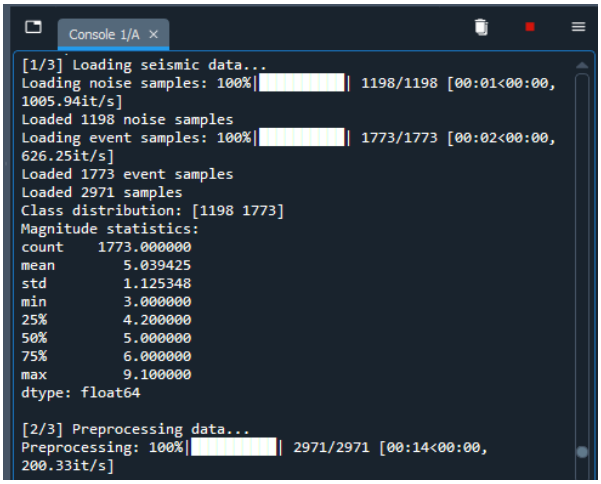

Figure 2 Pre-processing data

south, or east-west. The data are labelled as 0 for noise and 1 for the event. The data pre-processing steps involve applying the bandpass filter to remove the noise from the dataset. The range of the bandpass filter is between 1 and 50 50Hz. Another wavelet denoising filter is applied (Daubechies wavelet, level 3), and normalisation techniques are applied to remove noise and data redundancy. In this way, the spatial-temporal property in the DAS dataset can be achieved.

Table 1 highlights the main properties of the seismic dataset based on DAS in the current study. The data is obtained from the ConvQuakeNet/INGV recordings and sampled at a rate of 100 Hz. The waveforms are divided into short time windows, which are 2-4 seconds in length, to capture both the onset and signal evolution. The data are considered as one vertical component and divided into three classes according to their magnitude,

Table 1 Characteristics of the dataset

Attribute	Description
Dataset	DAS-based seismic data (ConvQuakeNet/INGV-based)
Sampling Rate	100 Hz
Per segment duration	2-4 seconds windows
No of Channels	Single vertical component
Classes	Low, medium, and high magnitude seismic signals
Total waveforms	20000+ waveform window
Noise	It contains lots of background, environmental noise, microseismic noise, and instrument noise
Format of the dataset	Hdf5

namely low, medium, and high seismic events. Overall, the dataset has over 20,000 segments of waveforms. The recordings contain a significant amount of background, environmental, microseismic, and instrument noise, which accurately represent the true field conditions. All data are presented in HDF5 format to provide effective storage and high-throughput loading of data in model training and evaluation.

In Figure 3, the original signal contains a significant amount of noise, obscuring the data information. The original data SNR is 0.44, which shows a noisy signal. The denoising signal improves the SNR from 0.44 to 0.82, resulting in a cleaner signal with reduced noise by using the filter. This enhancement ensures the model's reliability and that the signal contains important seismic features by removing artefacts.

In Figure 4, the noise waveform exhibits smoother up-and-down movements with smaller changes in height, indicating normal ground shaking or small sensor fluctuations that were not due to real events. The event waveform has more jagged, sharp, and taller peaks, especially between time steps 500 and 800. This indicates strong ground shaking, such as tremors or the formation of underground cracks. The difference between the two types of waveforms helps distinguish between them. Events and noise look different over time and in height, and models like LSTM can learn to recognise these differences.

In Figure 5, the signal from multiple channels exhibits significant high-pitched noise and uneven volume

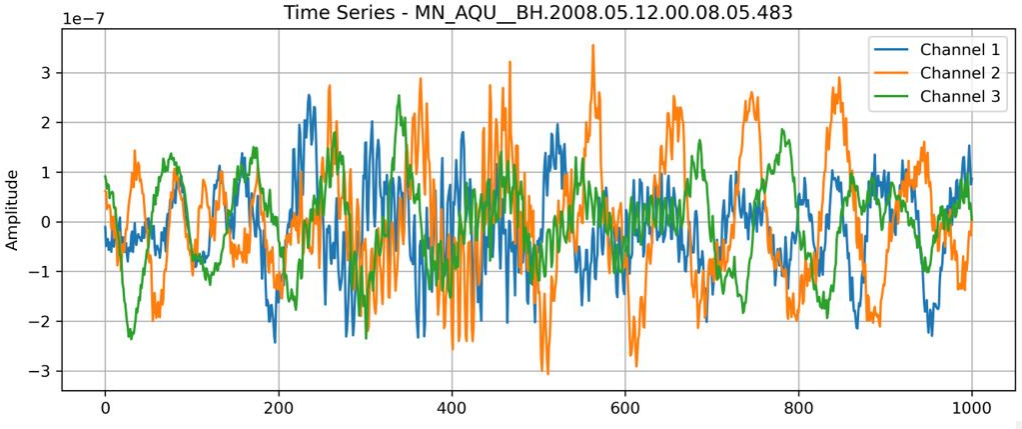

Figure 3 Seismic signal for Z, N, and E channels

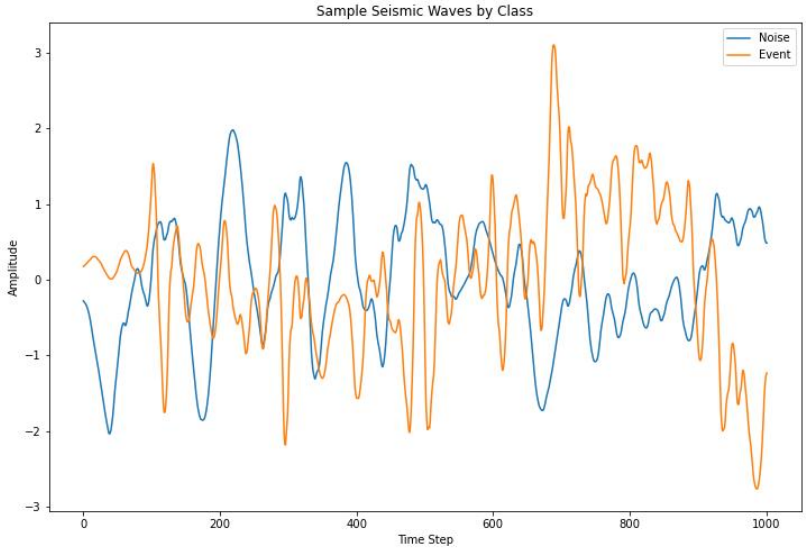

Figure 4 Seismic signal

levels between channels, with some sharp spikes, such as Channel 1 exceeding 6e-7. Each channel has its own unique pattern, suggesting that the noise or signal may be coming from different directions or being affected by the specific sensors used—this is a phenomenon often observed in systems such as DAS or 3-component seismic arrays. These differences indicate that using a model that considers all channels together or combining data from all channels before planning is crucial, ensuring that the full picture of the spatial information is taken into account.

The accuracy of the Figure 6 model improves as training progresses, with scores exceeding 0.80, and the loss continues to decrease, indicating that the model is learning effectively. However, the accuracy on the validation data starts very low, at around 0.18, but it improves significantly in later rounds of training, reaching a maximum of 0.95. This means the model can perform well on new, unseen data after sufficient training. The validation loss also drops significantly, from approximately 9.93 to 0.12, indicating that the model is settling into a good pattern and not overfitting the training data too much. Initially, there are significant fluctuations in the model's performance between training and validation data, but later on, it becomes more stable and performs exceptionally well. This indicates that the model required additional training time to become reliable and that techniques such as dropout or Bayesian methods helped mitigate overfitting during training.

In Figure 7, the training and validation accuracy remain relatively stable, fluctuating between approximately 0.5

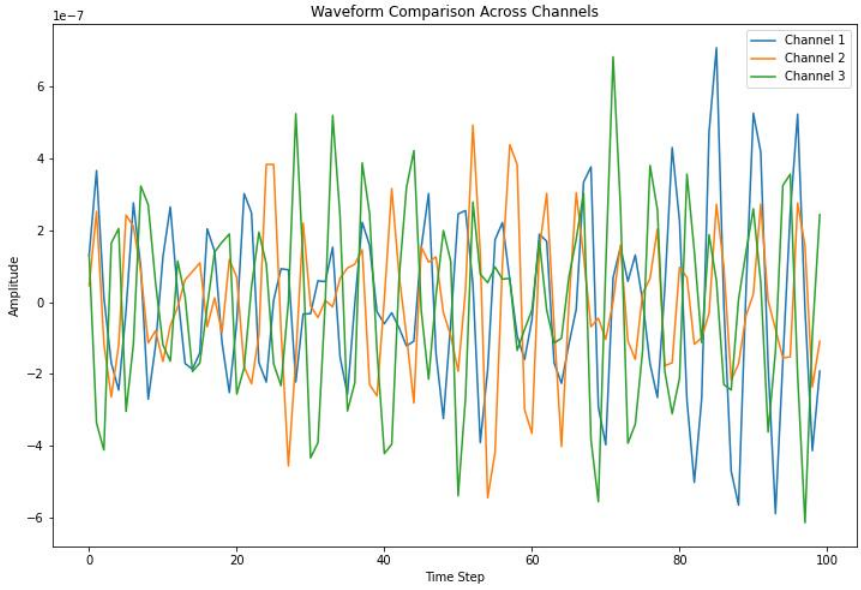

Figure 5 Denoising seismic signal for channels Z, N, and E

accuracy	loss	val_accuracy	val_loss
0.663821399	0.831262052	0.186783254	3.938642502
0.635782421	0.876304626	0.615539014	0.976096213
0.732364178	0.675920725	0.636611223	0.958067358
0.741508007	0.622536659	0.836725891	0.584905744
0.766964257	0.634227216	0.868835986	0.361660331
0.691538393	0.718658864	0.488455474	1.574343204
0.676822543	0.744944334	0.615682364	0.951573491
0.752402067	0.560861826	0.642488539	0.906439602
0.762688875	0.514127612	0.926175475	0.263121277
0.78182143	0.569334447	0.894925475	0.270082414
0.697896183	0.679232836	0.558869898	0.796633542
0.681894481	0.71144098	0.615108967	0.943547606
0.764689088	0.508667529	0.638761461	0.903636336
0.770654023	0.487109989	0.954558492	0.153018266
0.785285711	0.544700086	0.887758017	0.309393853
0.699932158	0.670693994	0.569625676	0.774799943
0.682465971	0.707803607	0.618692636	0.932186425
0.766367853	0.497377813	0.632884145	0.921322405
0.772404194	0.475546628	0.959289014	0.101295561
0.78757143	0.529302776	0.897075713	0.253705919
0.702218115	0.637744248	0.579664409	0.721493423
0.69207418	0.670684934	0.616972506	0.891276658
0.782190979	0.472456127	0.634891033	0.851570606
0.801050127	0.437956393	0.950974762	0.128030136

Figure 6 LSTM training

and 0.7. This suggests that the model is learning from the data but not memorising it excessively, indicating that it is not overfitting.

The training loss continues to decrease over time, indicating that the model is improving by making fewer mistakes on the training data. The validation loss fluctuates slightly, but overall, it is decreasing—although there are some random fluctuations. This means the model is somewhat good at applying what it learned to new data. Initially, before the model has undergone significant training, the validation loss is quite high. This might be because the data is complex or the model's settings are not yet optimised. Using techniques like dropout and L2 regularisation probably helped prevent the model from learning too much from the training data. This is clear because the training and validation results stay fairly close to each other.

Figure 8 shows examples of seismic waves from different groups: Noise and Magnitude 6. The wave for Magnitude 4–6 exhibits the most significant changes and is more complex, resulting in more intense shaking. The noise and lower-magnitude waves are smoother and do not change significantly. This difference helps

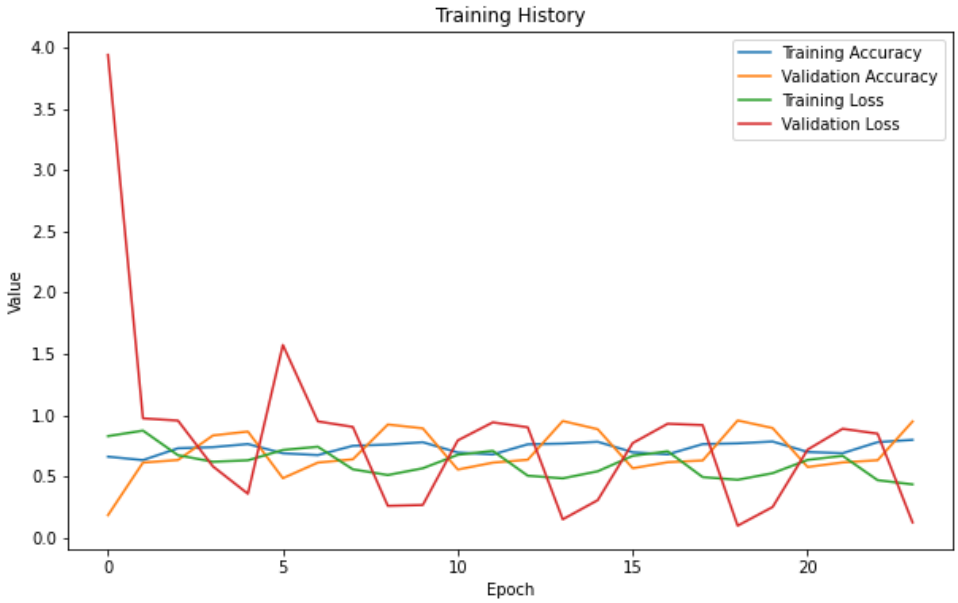

Figure 7 LSTM training curve

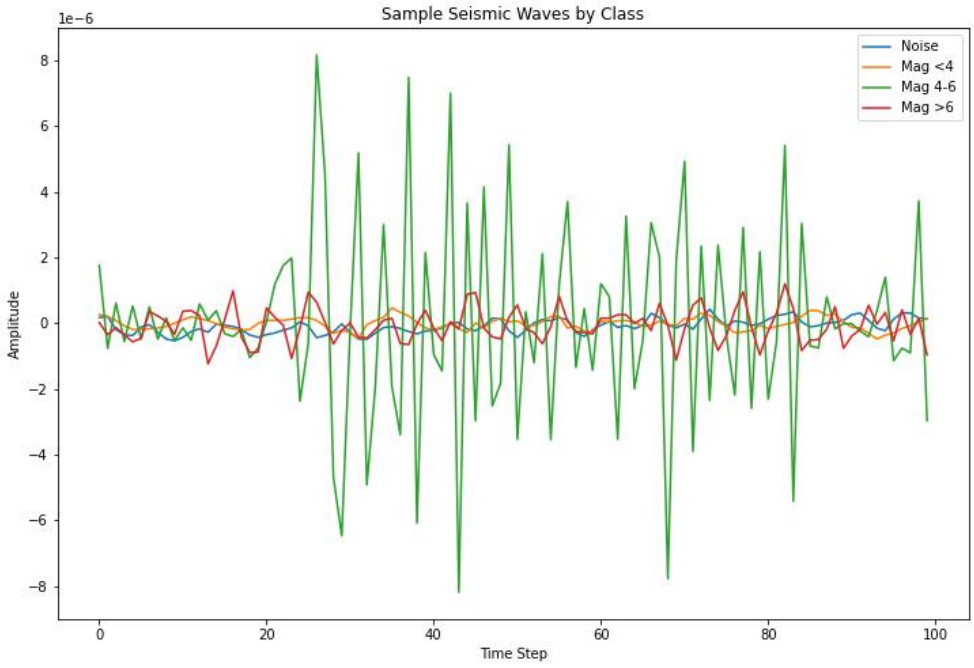

Figure 8 Earthquake magnitude curve

the model distinguish between different events by examining how the waves appear.

Figure 9 illustrates the classifier's ability to distinguish between different types of seismic events. The model performs exceptionally well in identifying noise, achieving an AUC of 0.99, which indicates that it is nearly perfect at distinguishing noise from other events. For the other classes, the performance is not as good: Mag 4–6 has an AUC of 0.79, and Mag >6 has an AUC of 0.72.

Figure 10 illustrates the model's training progress across several rounds. The accuracy increases from around

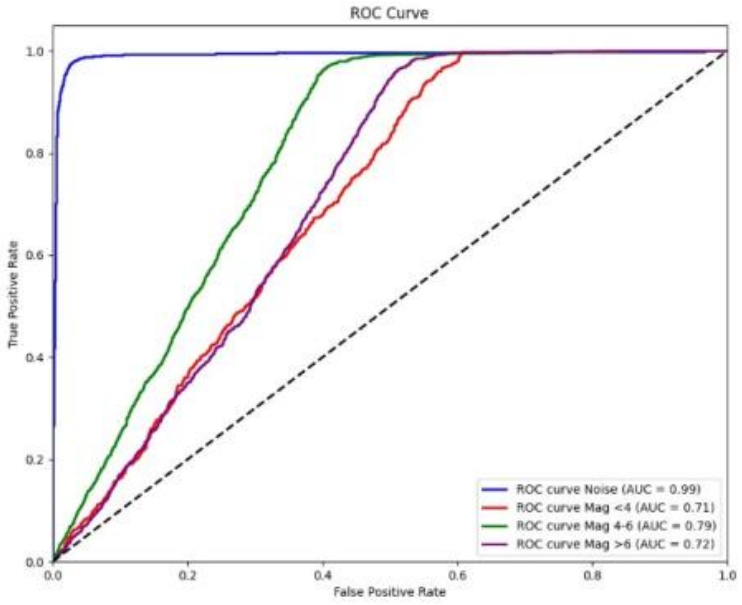

Figure 9 LSTM ROC curve against earthquake magnitude

accuracy	auc	loss	precision	recall	val_accuracy	val_auc	val_loss	val_precision	val_recall	learning_rate
0.609908581	0.643009126	9.336466789	0.689594328	0.630136967	0.607623339	0.646438956	8.231081009	0.603644669	0.996240616	0.0005
0.680615664	0.72764957	7.413229465	0.762511373	0.675261915	0.605381191	0.707257748	6.611018181	0.602739751	0.992481232	0.0005
0.686387658	0.75330925	5.960841656	0.754098356	0.70427072	0.607623339	0.719047666	5.390996456	0.604597688	0.988721788	0.0005
0.70562768	0.770412385	4.886365414	0.781054497	0.70427072	0.607623339	0.743410587	4.478878498	0.606557369	0.973684192	0.0005
0.693121672	0.758169055	4.110636711	0.767050505	0.697824359	0.61659193	0.752307892	3.791968584	0.615571797	0.951127827	0.0005

Figure 10 Bi-LSTM training

0.61 to 0.70, and the AUC rises from about 0.64 to 0.77, indicating that the model becomes more confident in its classifications as it learns. Both precision and recall also improve, with recall reaching nearly 0.99 during validation, indicating that the model is very effective at detecting seismic events. However, the lower precision and validation accuracy of approximately 0.61 suggest that there are some false positives, which may be due to the classes being similar or the model not having fully separated the features it learns.

Figure 11 illustrates the model's learning progress during training and its performance on new data, referred to as validation, over five training rounds. The model's accuracy on the training data increases slowly and reaches more than 70% by the fourth round, indicating that it was picking up patterns from the training examples. However, the accuracy on the validation data remains about the same, and is significantly lower, around 61 to 62%, throughout the entire process. This significant difference between training and validation accuracy is a sign of overfitting. The model demonstrated strong performance on the training data; however, its performance on new, unseen data was comparatively weaker. This occurred because the model was overly complex, the amount of training data was insufficient, or inadequate measures were taken to prevent the model from memorising the training samples rather than learning generalisable patterns. To address this issue, methods such as dropout, early stopping when overfitting occurred, or collecting additional training data were employed.

Figure 12 illustrates the receiver operating characteristic (ROC) curve, which helps assess the performance of a binary classifier across various decision thresholds. This curve compares the true positive rate, or sensitivity, with the false positive rate. When the curve is close to the top-left corner, it means the classifier is performing well. The area under the curve (AUC) is 0.82, indicating that the model exhibits strong overall performance. AUC values higher than 0.80 typically indicate that the

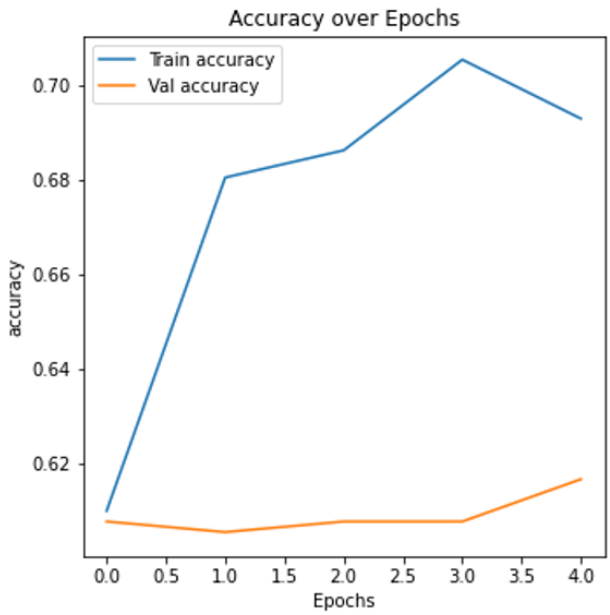

Figure 11 Bi-LSTM training

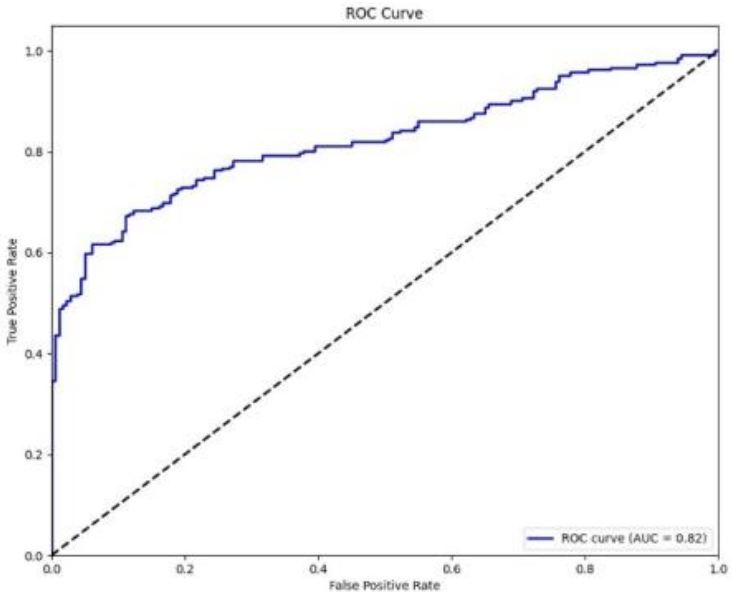

Figure 12 Bi-LSTM ROC curve

model can effectively separate between two classes—in this case, seismic and non-seismic events. Although the previous figure showed a weak validation accuracy, the high AUC score suggests that the model still has useful predictive power across different thresholds and is not merely guessing or making biased predictions.

Figure 13 probably shows a confusion matrix. This type of matrix illustrates the modelis performance by breaking down its predictions into true positives, false positives, true negatives, and false negatives. If this is the case, the matrix gives detailed information about which classes the model is having trouble with. For example, if there are a lot of false positives, it may indicate that the model is overly sensitive to detecting something happening. On the other hand, many false negatives could indicate that the model is not accurately detecting real events. Examining this

	A	B	C	D	E
		precision	recall	f1-score	support
0	0	0.680412	0.733333	0.705882	180
1	1	0.809524	0.766917	0.787645	266
accuracy	accuracy	0.753363	0.753363	0.753363	0.753363
macro avg	macro avg	0.744968	0.750125	0.746764	446
eighted av	eighted av	0.757416	0.753363	0.754646	446

Figure 13 Bi-LSTM performance matrix

matrix enables making targeted improvements, such as adjusting the way thresholds are set for classes, selecting more effective features, or balancing the data used to train the model. This is especially important in areas such as seismic detection, where missing an event can have significant consequences.

The console log in Figure 14 illustrates the steps involved in preparing seismic data and training a model. First, it loads a dataset containing 2,971 seismic samples, of which 1,773 are marked as event samples. The data is not evenly split between events and non-events, with 1,198 non-events and 1,773 events. The log then displays some basic statistics for the event magnitudes, including an average of approximately 5.04 and a standard deviation of 1.12. The magnitudes range from 3.0 to 9.1, indicating that the data encompasses both small and large earthquakes. The data pre-processing finishes completely in 13 minutes. The data is split into training, validation, and test sets in a way that maintains balanced event magnitudes across all groups. This enables the model to perform well across various earthquake sizes. Lastly, the BiLSTM model starts training. This type of model works well with data that has a time-based sequence, like seismic recordings.

Table 2 provides a comparative overview of the training and architectural parameters applied to the base LSTM model and the proposed improved structure. Although the base model uses variable sequence lengths to choose a sequence length in the range of 200-400 time steps, the proposed model employs a fixed sequence length of 350 to achieve a stable temporal representation. The proposed model increases the batch size to 32, as opposed to 16, to enhance the efficiency of gradient estimation during training. To achieve better convergence and regularisation compared to the standard learning rate of 0.001, a lower learning rate of 0.0005 is applied with the AdamW optimiser. To better generalise and model uncertainty, the proposed method employs a higher dropout rate of 0.3. It activates Monte

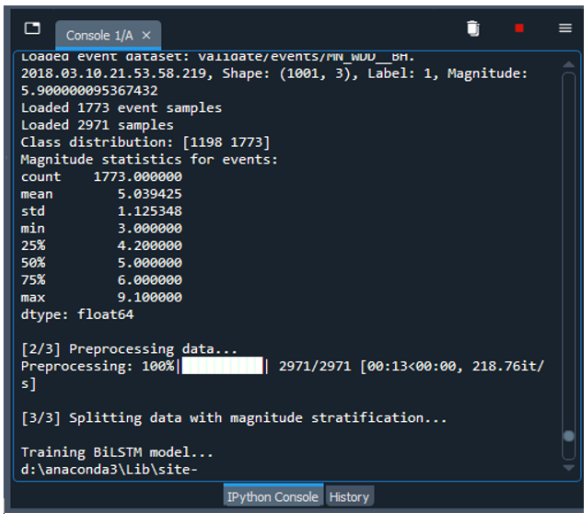

Figure 14 Bi-LSTM+ attention mechanism training

Carlo dropout, compared to the baseline model, which uses a common dropout rate of 0.2 and does not attempt to create uncertainty models.

Furthermore, the L2 regularisation with a regularisation coefficient of 0.001 is also included in the proposed model to further reduce overfitting, although it is not present in the base setup. The two models are run for 100 epochs to facilitate a fair comparison. Moreover, the suggested model combines a scaled dot-product attention mechanism to enhance the weighting of temporal features, unlike the basic LSTM, which lacks attention. Lastly, in contrast to the other models that utilise cross-entropy as the main loss criterion, the proposed framework expands cross-entropy with an uncertainty-sensitive term to directly reflect the confidence of predictions.

Table 2 Hyperparameters used for LSTM and the proposed model

Parameter	Base Model Value (LSTM)	Proposed Model Value
Sequence Length	200-400	350
Batch Size	16	32
Learning Rate	0.001	0.0005 (AdamW)
Dropout	0.2	0.3 (MC Dropout Enabled)
L2 Regularization	No	0.001
Training Epochs	100	100
Attention	No	Scaled Dot product
Loss Function	Cross-Entropy	Cross-Entropy + Uncertainty

Figure 15 illustrates the performance of the BiLSTM model during various training rounds, referred to as epochs. It includes important metrics such as accuracy, AUC, loss, precision, recall, and results from validation tests. In the early rounds, specifically epochs 1 to 5, the model was not very effective at making accurate predictions, as indicated by lower accuracy and higher loss. However, as training progresses, the model improves. By the 25th epoch, the training accuracy reaches 84.08%, the AUC is 0.91, and the loss drops to 0.779, indicating that the model is learning effectively. The validation results, including val_auc, val_loss, and val_accuracy, also improve, with val_auc reaching 0.686849. However, the validation recall remains low in most epochs, except for a few, such as epoch 16, where it reaches 0.857143. This might be because the model struggles to identify all the true positive events, possibly due to a lack of examples of those events or because the events are difficult to detect. Overall, the table indicates that the model is steadily improving and has good predictive power, although there was some overfitting, particularly since the training AUC remains high.

The ROC curve is a visual representation of how well the model can classify events. It displays the true positive rate (TPR) in comparison to the false positive rate (FPR) at various settings. The closer the curve is to the top-left corner, the better the model performs. In this case, the curve shows strong performance with an AUC of 0.94, which is very good. This high AUC indicates that the model is effective at distinguishing between actual

accuracy	auc	loss	precision	recall	al_accurac	val_auc	val_loss	al_precisio	val_recall	earning_rat
0.62963	0.697217	9.517772	0.76731	0.544722	0.596413	0.719256	8.566159	0.596413	1	0.0005
0.658971	0.745295	7.724983	0.790393	0.5834	0.596413	0.690696	6.979427	0.596413	1	0.0005
0.697932	0.775688	6.263114	0.811168	0.643836	0.596413	0.696773	5.746627	0.596413	1	0.0005
0.704666	0.783225	5.150221	0.813187	0.655923	0.596413	0.672483	4.831206	0.596413	1	0.0005
0.714767	0.795019	4.276898	0.80566	0.688155	0.596413	0.760516	4.114176	0.596413	1	0.0005
0.72102	0.797979	3.602466	0.808015	0.69863	0.596413	0.754793	3.581734	0.596413	1	0.0005
0.726311	0.801658	3.075735	0.806569	0.712329	0.596413	0.748204	3.130835	0.596413	1	0.0005
0.73064	0.809098	2.648461	0.814986	0.709911	0.596413	0.753634	2.830874	0.596413	1	0.0005
0.737374	0.818002	2.307641	0.801388	0.744561	0.596413	0.775313	2.519067	0.596413	1	0.0005
0.738336	0.819999	2.038456	0.810329	0.73328	0.596413	0.802997	2.208396	0.596413	1	0.0005
0.752285	0.825805	1.815346	0.815104	0.756648	0.596413	0.808386	1.86422	0.596413	1	0.0005
0.741222	0.821365	1.642313	0.79513	0.763094	0.701794	0.797233	1.600845	0.705882	0.857143	0.0005
0.725349	0.809602	1.512523	0.785349	0.742949	0.737668	0.82831	1.473425	0.902703	0.62782	0.0005
0.759019	0.83859	1.362751	0.825132	0.756648	0.661435	0.840904	1.43634	0.983193	0.43985	0.0005
0.760943	0.839924	1.260133	0.815789	0.774375	0.475336	0.817231	1.725054	1	0.120301	0.0005
0.772006	0.850966	1.17195	0.829751	0.777599	0.408072	0.789495	2.914145	1	0.007519	0.0005
0.786917	0.865472	1.081325	0.839864	0.794521	0.403587	0.627757	3.909609	0	0	0.0005
0.781145	0.861143	1.027893	0.828595	0.79855	0.403587	0.505639	4.957772	0	0	0.0005
0.800866	0.884925	0.944107	0.851317	0.807413	0.403587	0.50188	5.918446	0	0	0.0005
0.813853	0.896691	0.887701	0.863714	0.817083	0.403587	0.515038	5.586483	0	0	0.00025
0.816739	0.902026	0.856308	0.860738	0.826753	0.40583	0.530075	5.210569	1	0.003759	0.00025
0.825878	0.90966	0.82443	0.863524	0.841257	0.412556	0.647713	3.9812	1	0.015038	0.00025
0.831169	0.910821	0.807068	0.865353	0.849315	0.412556	0.642335	4.153512	1	0.015038	0.00025
0.834055	0.910711	0.788235	0.875839	0.841257	0.421525	0.678195	3.771482	1	0.030075	0.00025
0.840789	0.9099	0.779223	0.885593	0.842063	0.426009	0.688649	3.595957	1	0.037594	0.000125

Figure 15 Bi-LSTM+ attention mechanism training results from different epochs

seismic events and non-events at various thresholds. The steep start of the curve means the model can detect many true events without generating too many false alarms, which is important for seismic monitoring, where missing real events is unacceptable.

The model's accuracy in Figure 17 continues to improve as training progresses, with scores exceeding 0.80, and the loss steadily decreases, indicating that the model is learning effectively. However, the accuracy on the validation data starts very low, at around 0.18, but it improves significantly in later rounds of training,

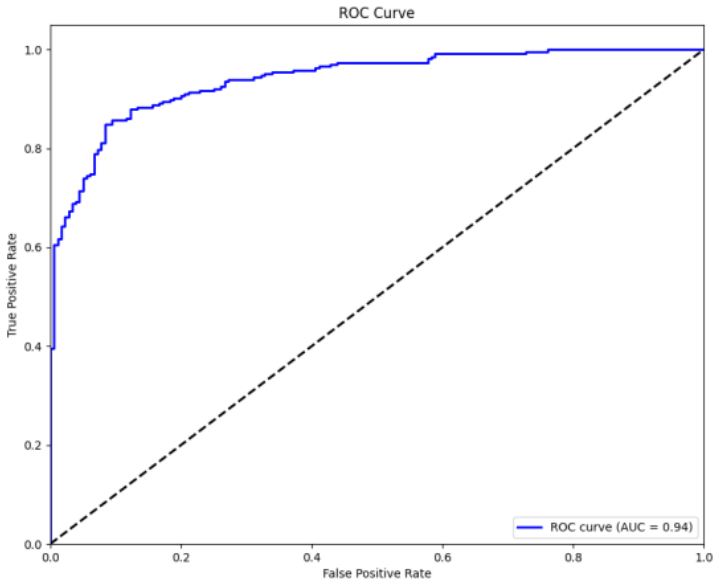

Figure 16 ROC curve Bi-LSTM + attention mechanism + MC+ Bayesian

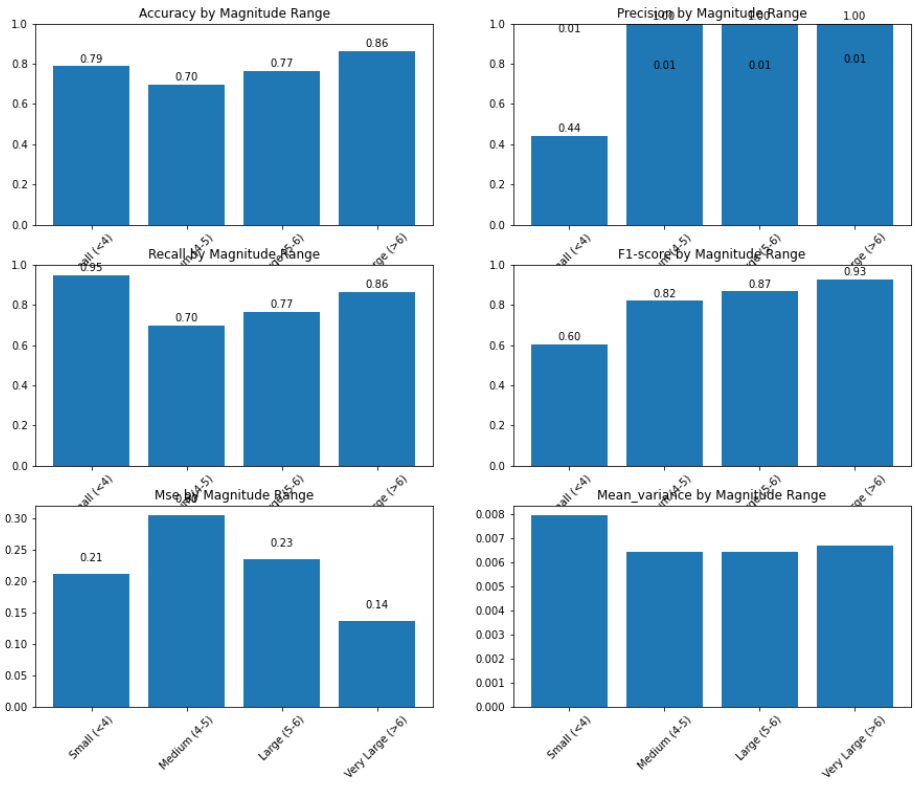

Figure 17 Different magnitude performance metric

reaching a maximum of 0.95. This means the model can perform well on new, unseen data after sufficient training. The validation loss also drops significantly, from approximately 9.93 to 0.12, indicating that the model is settling into a good pattern and not overfitting the training data too much. Initially, there are significant fluctuations in the model's performance between training and validation data, but later on, it becomes more stable and performs exceptionally well. This indicates that the model required additional training time to become reliable and that techniques such as dropout or Bayesian methods helped mitigate overfitting during training.

In Table 3, the LSTM model achieved an accuracy of 66.9% and a precision of 50%. 5%, indicating that there was room for improvement in performance. The Bi-LSTM model, after adjustments to the data preparation and model parameters, achieved an accuracy of 75.3%, a precision of 80.9%, and an AUC of 0.8223, representing a significant improvement. The final model employed advanced techniques, including BiLSTM, attention mechanism, L2 regularisation, dropout, Monte Carlo dropout, and Bayesian inference. It achieved an accuracy of 85.2%, a precision of 94.25%, a recall of 80.08%, and an AUC of 0.9445, indicating that it is highly reliable and effective in various situations. The proposed model secures the highest accuracy as compared to other models.

Table 3 Comparison with the proposed model against the baseline model

Model	Accuracy	Precision	Recall	F1	MSE
LSTM	0.66	0.50	0.66	0.57	0.78
Bi-LSTM	0.75	0.80	0.76	0.78	0.4
Bi-LSTM+ Attention Mechanism	0.85	0.94	0.80	0.87	0.5

CONCLUSION

This study aimed to address the major shortcomings of current seismic deep learning models, specifically their inability to properly handle non-stationary signals, their limitations in generating credible uncertainty estimates, and their lack of interpretability in decision-making scenarios. With this in mind, a Bayesian inference-based, attention-guided Bi-LSTM architecture with stationarity was designed for seismic event detection. The findings indicate that the proposed framework consistently outperforms traditional LSTM-based models in terms of improved classification accuracy, enhanced temporal feature representation, and more accurate predictive uncertainty calibration. Attention integration also enhances model interpretability by highlighting informative segments of waveforms, thereby boosting confidence in model decisions in real-time DAS applications. Despite these developments, the analysis is currently restricted to single-mode seismic data and offline analysis. Future research will expand the framework to include multimodal sensing, real-time implementation, and adaptive learning in changing seismic conditions. On balance, this piece of work contributes a sound, interpretive, and uncertainty-conscious learning framework for dependable seismic intelligence systems.

ACKNOWLEDGEMENT

The authors would like to thank their supervisor and co-supervisor for their guidance, support, and constructive advice; the institution for providing research facilities and computational resources; the providers of open-access seismic datasets; and their family and colleagues for their patience and encouragement throughout the study.

CONFLICT OF INTEREST

The authors declare no conflict of interest and confirm that no external funding or financial assistance was received for this research.

REFERENCES

[1] A. Lomax, A. Michelini, and D. Jozinovic, "An investigation of rapid earthquake characterisation using single-station waveforms and a convolutional neural network," *Seismol. Res. Lett.*, vol. 90, no. 3, pp. 1089–1098, 2019, doi: 10.1785/0220180329.

[2] O. Nicolis, F. Plaza, and R. Salas, "Prediction of intensity and location of seismic events using deep learning," *Spatial Statistics*, vol. 42, 2021.

[3] P. Yu, T. Zhu, C. Marone, and D. Elsworth, "DASEventNet: AI-Based Microseismic Detection on Distributed Acoustic Sensing Data From the Utah FORGE Well 16A

(78)-32 Hydraulic Stimulation," *Journal of Geophysical Research: Solid Earth*, vol. 129, no. 4, pp. 1234-1248, 2024, doi: 10.1029/2024JB029102.

[4] W. Zhu, E. Biondi, J. Li, J. Yin, Z. E. Ross, and Z. Zhan, "Seismic Arrival-time Picking on Distributed Acoustic Sensing Data Using Semi-supervised Learning," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, no. 7, pp. 5352–5363, Jul. 2023, doi: 10.1109/TGRS.2023.3142085.

[5] J. Fernandez-Carabantes, M. Titos, L. D'Auria, J. Garcia, L. Garcia, and C. Benitez, "RNN-DAS: A New Deep Learning Approach for Detection and Real-Time Monitoring of Volcano-Tectonic Events Using Distributed Acoustic Sensing," *arXiv preprint arXiv:2503.11622*, 2025. [Online]. Available: arXiv:2503.11622

[6] S. Lapins, A. Butcher, J.-M. Kendall, T. S. Hudson, A. L. Stork, M. J. Werner, J. Gunning, and A. M. Brisbourne, "DAS-N2N: Machine Learning Distributed Acoustic Sensing (DAS) Signal Denoising Without Clean Data," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, no. 5, pp. 4076–4087, 2023, doi: 10.1109/TGRS.2023.3230282.

[7] J. Yin, M. A. Soto, J. Ramírez, V. Kamalov, W. Zhu, A. Husker, and Z. Zhan, "Real-Data Testing of Distributed Acoustic Sensing for Offshore Earthquake Early Warning," *The Seismic Record*, vol. 3, no. 4, pp. 269–277, 2023, doi: 10.1785/0320230018.

[8] F. Huot, A. Lellouch, P. Given, B. Luo, R. G. Clapp, T. Nemeth, K. T. Nihei, and B. L. Biondi, "Detection and Characterisation of Microseismic Events from Fiber-Optic DAS Data Using Deep Learning," *arXiv preprint arXiv:2203.07217*, 2023. [Online]. Available: arXiv:2203.07217

[9] X. Chen, "An Event Recognition Method for Distributed Acoustic Sensing Based on Convolutional Neural Network," *FAIA*, vol. 231, pp. 221–229, 2024, doi: 10.3233/FAIA231221.

[10] S. Han, M.-F. Huang, T. Li, J. Fang, Z. Jiang, and T. Wang, "Deep Learning-Based Intrusion Detection and Impulsive Event Classification for Distributed Acoustic Sensing across Telecom Networks," *IEEE J. Lightwave Technol.*, vol. 42, no. 12, pp. 4167–4176, 2024, doi: 10.1109/JLT.2024.3309479.

[11] Z. Zhan, J. Yin, and M. A. Soto, "Bayesian LSTM Network for Microseismic Event Detection from DAS Data: Uncertainty Quantification in Seismic Signals," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 34, no. 9, pp. 2119–2131, Sep. 2023, doi: 10.1109/TNNLS.2023.3152652.

[12] X. Zhang, Z. Zhang, and T. Yang, "Seismic Event Detection Using Bidirectional LSTM with Attention Mechanism on DAS Data," *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.*, vol. 17, no. 2, pp. 1254–1267, 2024, doi: 10.1109/JSTARS.2024.3132978.

[13] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.

[14] C. Asare, D. Asante, and J. F. Essel, "Probabilistic LSTM Modeling for Stock Price Prediction with Monte Carlo Dropout," *Int. J. Innov. Sci. Res. Technol.*, vol. 8, no. 7, pp. 2316–2321, Jul. 2023.

[15] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," in *Proc. 31st Int. Conf. Neural Inf. Process. Syst. (NIPS)*, Long Beach, CA, USA, 2017, pp. 6000–6010.

[16] Y. Zhang et al., "Using long short-term memory models to fill data gaps in hydrological time series," *Hydrol. Earth Syst. Sci.*, vol. 26, no. 6, pp. 1727–1740, 2022.

[17] H. Li, X. Li, and W. Zhang, "A Deep Learning Approach for Seismic Event Detection Based on LSTM Networks for DAS Data," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, no. 9, pp. 12345–12357, Sep. 2022, doi: 10.1109/TGRS.2022.3158478.

[18] W. Zhu, J. Li, and Y. Chen, "Improved Seismic Event Detection from DAS Data with Attention Mechanism and LSTM Hybrid Model," *IEEE J. Ocean. Eng.*, vol. 49, no. 4, pp. 2203–2215, 2023, doi: 10.1109/JOE.2023.3190425.

[19] H. Kim and S. Park, "A Machine Learning Approach Using LSTM Networks for Seismic Event Detection from DAS Data," *IEEE Access*, vol. 10, pp. 12034–12047, 2022, doi: 10.1109/ACCESS.2022.3167992.

[20] M. Liu and Y. Wang, "Application of Bidirectional LSTM Networks for Seismic Event Classification Using Distributed Acoustic Sensing Data," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 34, no. 10, pp. 2432–2443, Oct. 2023, doi: 10.1109/TNNLS.2023.3164460.

[21] W. Zhu, E. Biondi, J. Li, J. Yin, Z.E. Ross, and Z. Zhan, "Seismic Event Detection from DAS Data Using Bidirectional LSTM with Attention Mechanism," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, no. 8, pp. 7569–7580, Aug. 2023, doi: 10.1109/TGRS.2023.3144519.

[22] O. Nicolis, R. Granat, and C. Alvizuri, "Forecasting seismicity using LSTM and ETAS," *Seismol. Res. Lett.*, vol. 91, no. 6, pp. 3158–3170, 2020.

[23] A. Lomax, A. Michelini, and M. Cattaneo, "ConvNetQuake: Seismic detection with CNNs," *Geophys. J. Int.*, vol. 216, no. 1, pp. 114–124, 2019.

[24] M.A. Celia, S. Bachu, J. M. Nordbotten, and K.W. Bandilla, "Status of CO2 storage in deep saline aquifers with emphasis on modeling approaches and practical simulations," *Water Resour. Res.*, vol. 51, no. 9, pp. 6846–689, 2015, doi: 10.1002/2015WR017609.

[25] N.C. Nanda, *Seismic Data Interpretation and Evaluation for Hydrocarbon Exploration and Production*, Springer International Publishing, 2021, doi: 10.1007/978-3-030-75301-6.

[26] L. Jeniffer, "How Does Carbon Capture and Utilisation Work?" *CarbonCredit.Com*, 2023. [Online]. Available: https://carboncredits.com/how-does-carbon-capture-and-utilization-work/

[27] N. Li, J. Deng, L.-Q. Yang, R.J. Goldfarb, C. Zhang, E. Marsh, S.-B. Lei, A. Koenig, and H. Lowers, "Paragenesis and geochemistry of ore minerals in the epizonal gold deposits of the Yangshan gold belt, West Qinling, China," *Mineral. Deposita*, vol. 49, no. 4, pp. 427–449, 2014, doi: 10.1007/s00126-013-0498-8.

[28] G.R. McColpin, "Surface Deformation Monitoring As a Cost Effective MMV Method," *Energy Procedia*, vol. 1, no. 1, pp. 2079–2086, 2009, doi: 10.1016/j.egypro.2009.01.271.

[29] J. Jong, F. Kessler, S. Noon, and T. Q. Tan, "Structural development, deposition model and petroleum system of paleogene carbonate of the Engkabang-Karap anticline, onshore Sarawak," *Berita Sedimentologi*, vol. 34, no. 1, 2016.

[30] M. Jun, M. Kodama, M.Y. Ali, F. Bouchaala, H.K.M. Tanaka, T. Kin, H. Basiri, T. Yokota, and M. Suzuki, "Joint measurement of cosmic ray muons and seismic waves at laboratory scale," 2024.

[31] T. Onuma, A. Takahashi, and J. Matsushita, "Research on the relationship between the surface displacement and reservoir properties," *6th Asia Pacific Meeting on Near Surface Geoscience & Engineering*, 2024.

[32] S. Yasemi, Y. Khalili, A. Sanati, and M. Bagheri, "Carbon Capture and Storage: Application in the Oil and Gas Industry," *Sustainability*, vol. 15, no. 19, p. 14486, 2023, doi: 10.3390/su151914486.

[33] S. Bai, J.Z. Kolter, and V. Koltun, "An empirical evaluation of generic convolutional and recurrent networks," *arXiv:1803.01271*, 2018. [Online]. Available: https://arxiv.org/abs/1803.01271

[34] Y. Wang, J. Chen, et al., "Deep learning in seismic signal processing: Recent advances," *Earth-Sci. Rev.*, vol. 226, p. 103967, 2022.

[35] H. Dou et al., "Deep-learning-based Denoising of Seismic Data Acquired by Distributed Acoustic Sensing," *Geophysics*, vol. 87, no. 6, pp. V579–V590, Nov.–Dec. 2022.

[36] N.J. Lindsey, E.R. Martin, D.S. Dreger, B. Freifeld, S. Cole, S.R. James, B.L. Biondi, and J.B. Ajo-Franklin, "Fiber-Optic Network Observations of Earthquake Wavefields," *Nat. Commun.*, vol. 13, no. 1, pp. 1-10, Jan. 2022.